

Title (en)

Systems and methods for extraction of concepts for reuse-based schema matching

Title (de)

Systeme und Verfahren zur Extraktion von Konzepten für wiederverwendungs-basierten Schemaabgleich

Title (fr)

Systèmes et procédés pour l'extraction de concepts pour correspondance de schéma basée sur la réutilisation

Publication

EP 2600258 A3 20130731 (EN)

Application

EP 12008014 A 20121129

Priority

US 201113310483 A 20111202

Abstract (en)

[origin: EP2600258A2] In one embodiment, an approach to automated recurring concept extraction, from a plurality of input data models (schemas) is presented. The approach converts input data models to graphs, with typed elements. The graphs are mined for closed subgraphs that have a defined minimum support. The identified subgraphs can be filtered with a relevance metric. These subgraphs are converted to schemas or an appropriate representation, and stored for reuse in a repository. The repository can be used to automate further transformation or mapping of schemas presented to a system that uses the repository. In one example, the repository is used in a schema covering process to perform schema transformation.

IPC 8 full level

G06F 17/30 (2006.01)

CPC (source: EP US)

G06F 16/213 (2018.12 - EP US); **G06F 16/25** (2018.12 - EP US)

Citation (search report)

- [XY] LIU, YUBAO, FENG, YUCAI: "Interactive Mining of Schema for Semistructured Data", PROC. SPIE 4730, DATA MINING AND KNOWLEDGE DISCOVERY: THEORY, TOOLS, AND TECHNOLOGY IV, 13 March 2002 (2002-03-13), Orlando, Florida, pages 432 - 441, XP055067079, Retrieved from the Internet <URL:http://proceedings.spiedigitallibrary.org/data/Conferences/SPIEP/31356/432_1.pdf> [retrieved on 20130618], DOI: 10.1117/12.460250
- [IJ] BARNAS SAHA ET AL: "Schema covering: a step towards enabling reuse in information integration", DATA ENGINEERING (ICDE), 2010 IEEE 26TH INTERNATIONAL CONFERENCE ON, IEEE, PISCATAWAY, NJ, USA, 1 March 2010 (2010-03-01), pages 285 - 296, XP031657867, ISBN: 978-1-4244-5445-7
- [IY] "Mining Frequent Itemsets Using Support Constraints", PROCEEDINGS OF THE 26TH VLDB CONFERENCE ON VERY LARGE DATA BASES, 10 September 2000 (2000-09-10), Cairo, Egypt, pages 43 - 52, XP055067071, Retrieved from the Internet <URL:http://citeseerx.ist.psu.edu/viewdoc/download?sessionid=3C4CD74D1C6707E6AA4490C8646ED3B5?doi=10.1.1.37.5880&rep=rep1&type=pdf> [retrieved on 20130618]
- [YA] ERHARD RAHM ET AL: "Matching Large XML Schemas", vol. 33, no. 4, 4 December 2004 (2004-12-04), pages 26 - 31, XP002619063, ISSN: 0163-5808, Retrieved from the Internet <URL:http://delivery.acm.org/10.1145/1050000/1041415/p26-rahm.pdf?key1=1041415&key2=0308026921&coll=DL&dl=ACM&CFID=7038453&CFTOKEN=30493657> [retrieved on 20110126], DOI: 10.1145/1041410.1041415
- [YA] DO ET AL: "Matching large schemas: Approaches and evaluation", INFORMATION SYSTEMS, PERGAMON PRESS, OXFORD, GB, vol. 32, no. 6, 5 May 2007 (2007-05-05), pages 857 - 885, XP022062984, ISSN: 0306-4379, DOI: 10.1016/J.IS.2006.09.002
- [YA] "VLDB '02: Proceedings of the 28th International Conference on Very Large Databases", 20 August 2002, ELSEVIER, ISBN: 978-1-55-860869-6, article H DO ET AL: "COMA - A system for flexible combination of schema matching approaches", pages: 610 - 621, XP055010168, DOI: 10.1016/B978-155860869-6/50060-3
- [A] YAN: "CloseGraph: Mining Closed Frequent Graph Patterns", 24 August 2003 (2003-08-24), XP055065888, Retrieved from the Internet <URL:http://delivery.acm.org/10.1145/960000/956784/p286-yan.pdf?ip=145.64.134.240&acc=ACTIVE SERVICE&key=986B26D8D17D60C88D75A192E3112143&CFID=224510562&CFTOKEN=66001813&__acm__=1370850242_53e08caf2976bc8adb55c87> [retrieved on 20130610]

Designated contracting state (EPC)

AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

Designated extension state (EPC)

BA ME

DOCDB simple family (publication)

EP 2600258 A2 20130605; EP 2600258 A3 20130731; US 2013144893 A1 20130606; US 8719299 B2 20140506

DOCDB simple family (application)

EP 12008014 A 20121129; US 201113310483 A 20111202